Sign Language to Text Conversion – A Survey

Adithi Krishnan[1], Ruthvik B R [1], Spoorthy M [1], Rhea Muthanna [1], Shashank N[2]

[1, 2] Department of Computer Science and Engineering
Vidyavardhaka College of Engineering, Mysuru, Karnataka India
[1]ruthvikbr24@gmail.com
[2]shashank.n@vvce.ac.in

Abstract. Sign languages are languages that use the visual-manual modality to convey meaning. It is a communication system using gestures usually used by deaf and dumb people to communicate with each other and with other people. As sign language is not a common one that people know, only people well versed in sign language are able to interpret the gestures and communicate with the mute. Hence, a need arises to bridge this gap and this is our aim. We plan to develop a mobile application that reads in sign language and converts it to text that many people understand. Most of the techniques present now have few disadvantages like less accuracy, skin tones, motion gestures, clutter, variability etc. Our main aim will is to develop a mobile application that converts sign language to text while also trying to mitigate the above drawbacks to some extent.

Keywords: Sign to Text Conversion, Convolutional neural networks, Deep learning, Android.

1 Introduction

Sign language detection and conversion is a multi-step process that includes object detection, image processing and feature extraction. Object detection is a computer technology for computer vision that detects objects such as hand signs, faces etc. Image Processing is a method that performs operations on a given image to enhance it or extract information useful to us. Once object is detected, we apply image processing techniques to remove noise/clutter and obtain a simplified version of the image. Feature extraction is a process by which we obtain relevant information from data and represent it in lower dimensionality space. Once we get the enhanced image, we apply feature extraction techniques on it to get useful information.

Once feature extraction is done, we use this relevant information to train a deep learning model. Deep learning is a branch of Machine Learning and Artificial Intelligence where we can train a network on unsupervised data. These are usually neural networks and can be a Convolutional Neural Network, Recurrent Neural

Network, Generative Adversarial Networks etc. A CNN would be most suitable in the case of converting sign language to text as it is a type of neural network that uses layers of perceptrons to analyze data and train a model. It applies to image processing, Natural Language Processing and many other tasks related to cognitive capabilities.

Once a trained model is obtained, it is deployed on the cloud and then an interface has to be developed to use that model to detect and understand gestures. We would also need a camera to record the gestures and an output device to display the text. Most people use smartphones and hence developing a mobile application is a very convenient way to achieve this. A mobile application is developed where we record the sign language gestures and pass it on to the model on the cloud for conversion. The resultant text obtained can then be displayed on the screen.

The proposed system would work like this:

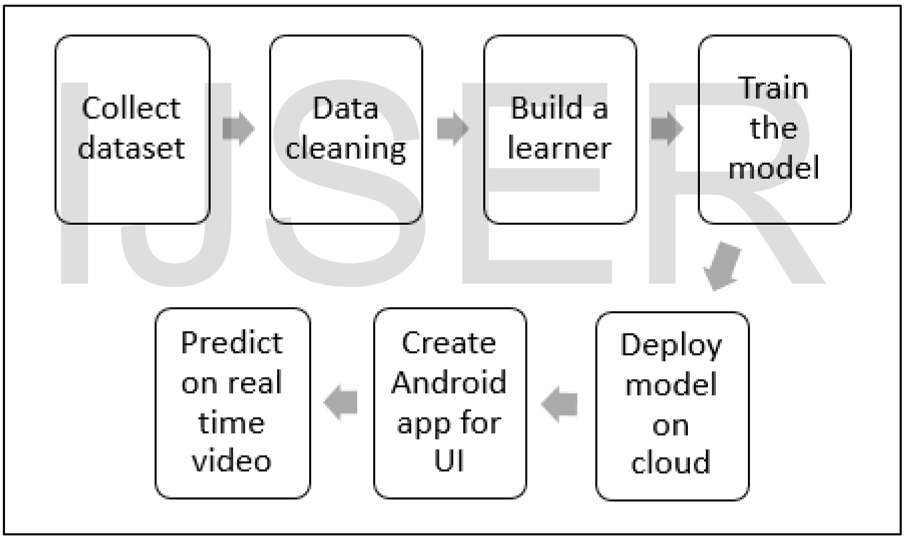

Fig 1: The proposed system to convert sign language to text.

2 Related Works

Bantupalli et al., [1] have used a convolutional neural network for spatial feature extraction, long short term Memory and recurrent neural networks for temporal feature extraction and then used ADAM optimizer and a softmax layer for prediction. Modi et al., [2] have developed a system that obtains frames from a video every 4 seconds and matches it with database to find out least error. Abdulla et al., [3] have used sensors that have RF transmitters that send signals based on hand movement to a receiver that translates it into Arabic language. Dutta et al., [4] have developed a system that calculates Eigen values for images and the takes pre-processed image as input and checks for maximum match.

Padmavathi et al., [5] have developed a system which Convert image to frames and apply HSI color model based segmentation on each. Neural networks are used to predict the character. Anand et al., [6] have developed a system which creates image feature vector by binarization, noise removal and hand detection in image and then compare with existing database. For speech to sign conversion, remove noise from audio, convert to text and then compare with existing database. Ong et al., [7] have developed a system which detects hand signs, hand shape and position of the hand across all positions and scales in the given image after removing erroneous values. Bhat et al., [8] have developed a system which uses Sensors in gloves to pick up gestures, convert to text with Analog to Digital Converter and microcontrollers, sends it to phone via Bluetooth which then converts text to speech.

Pramada et al., [9] have developed a system which captures image and performs RGB color detection and converts it to binary image and performs pattern matching and text to speech conversion. Huang et al., [10] have developed a system which uses Microsoft Kinect as input device and gives color and depth video streams and uses five inputs. The CNN has 9 frames and 8 layers, subsampling and convolution is done multiple times. Madhuri et al., [11] have developed a system in which image is obtained using a mobile camera and then image processing is done to extract hand sign and match it and then corresponding audio file is played. Wu et al., [12] have developed a system in which classifier is trained on both positive and negative data, weak classifier with lowest error rate is chosen in each run and then all the classifiers are combined as a strong classifier to detect the meaning of the gestures.

Jarndal et al., [13] have developed two systems for conversion to dual language (English and Arabic) text and voice, a vision based system and a wireless-interfaced glove based system. Hays et al., [14] have developed a mobile application for real time sign language to text conversion from a video input using classification algorithms Locality Preserving Projections (LPP) and Support Vector Machine (SVM). Vijayalakshmi et al., [15] have developed a flex sensor, tactile sensor and accelerator, HMM based sign language to text and speech conversion model.

3 Comparison of different sign language translation methods

The following table 1 gives us an idea of the different methods used in the field of sign language detection and translation by different authors. It also illustrates some of our recommendations that we thought could be implemented.

Table 1. Comparison of different methods used for sign language conversion

Objective	Methodology	Results/ Outcome	Advantage	Disadvantage
Create a vision based application that offers sign language translation to text, aiding communication.	A CNN named 'Inception' is used in spatial feature extraction from the video, then using a LSTM and a RNN model we extract temporal features using the outputs of softmax and the pool layer of CNN.	Accuracy was more for softmax layer than pool layer for various sample sizes.	As CNN and RNN are trained independently there is a minimization of cross-entropy-cost function ADAM.	This model faced problems while testing with different skin tones. It dropped accuracy if it hadn't been trained on a certain skin tone.
A method to enable translating Sign language finger-spellings to English text and enable finger-spelling to Digital ,Audio or Text conversion.	Extraction of video every 4 seconds and processing it, extracted features are compared with database of finger-spellings. Error is calculated, one with minimum error is the best match.	This approach results in the clear comparison obtained for each finger-spelling with all other database images. It resulted in 96% accuracy.	Simple mechanism to detect the finger-spelling. Easy to implement with the desired output probability of 0.96.	There is a need of addition of more gestures and features so that it supports motion too.
To develop a device which uses gloves to convert sign language to Arabic text language.	Five flex sensors detect the binding of each finger. The Arduino NANO interfaced with RF transmitter transmits the signals to receiver, The received signals generates Arabic letters and are displayed on LCD screen.	When a person wearing the smart gloves does an Arabic letter gesture, the LCD displays the letter and the speaker outputs the voice when sound button is clicked.	Low cost and it can be used to represent a wider range of words.	There is a need of combining two gloves instead of one as it doesn't cover wider range of sign and use of smart gloves is a compulsion.
To develop a system trained to convert sign language to text language using single and double hand and Min Eigen value	Min Eigen value is applied on 5 images of each alphabet and preprocessed input image. Interesting points are extracted. Check for max matching.	The test image and the database image is compared by matching the feature points and the database image matching is displayed as text and later to speech.	It is carried out with bare hands and the results were background and person independent.	
To convert the Indian sign language hand gestures to appropriate text message.	Convert image to frames and apply HSI color model based segmentation on each. Features line centroid of the hand is extracted and is fed to neural network to recognize the particular character.	The accuracy percentage of the result obtained is 99%,precision 89.47%,recall 89.78% and specificity 97.54%	Accuracy is more. This approach gave better results with sigmoid transfer.	Improper segmentation which results in varying of robustness when hands are overlapped.
Ease communication between deaf/dumb and normal people without use of sophisticated devices like data gloves etc.	Create image feature vector, noise removal, hand detection in image, compare with existing database. For speech to sign conversion, remove noise from audio, convert to text, and compare with existing database.	Not implemented yet.	Convenient way of communication between deaf/dumb and normal people with two way translation.	Should be extended to words and sentences. Difficult to implement image processing techniques on mobile phones.
Training a detector to recognize human hand in the image and also classify the hand shape.	Exhaustive detection across all positions and scales, thresholding image, connected component analysis to detect position of hand. Sub image in area of detected hand given to hand shape detectors.	99.8% success rate on hand detection and 97.4% success rate in hand shape classification.	Unsupervised approach trained using K-mediod algorithm, very efficient on grey-level images.	No motion or background mod els, accuracy not evaluated in envi ronments with more clutter and va iability.
Improve communication in Indian Sign Language using flex sensor technology.	Sensors in gloves pick up gestures, convert to text with ADC, microcontrollers and send to phone via Bluetooth which then converts text to speech.	Successfully converts Indian Sign Language, numbers and symbols to text and displays it on mobile phone.	Reliable, user independent and portable system and consumes less power compared to other	

Ref. No	Concept Used	Results/ Outcome	Advantage	Disadvantage
#1	Capture image, RGB color detection and conversion to binary image, thresholding, coordinate mapping, pattern matching and text to speech conversion.	Successfully recognizes the alphabet using Binary Finger-Tapping code.	Reliable and feasible since an inexpensive computer is introduced in the communication path.	It is challenging to recognize signs that involve motion.
#2	Microsoft Kinect as input device, give color, depth video streams, five inputs, CNN has 9 frames and 8 layers, subsampling and convolution multiple times.	3D CNN with gray image has higher accuracy than 3D ConvNet and for multi channels 3D CNN method outperforms other methods	3D CNN method outperforms other methods	
#3	Image is obtained using a mobile camera and then image processing to extract hand sign and then recognition stage , corresponding audio file is played	This sign language translator is able to translate Alphabets (A-Z) and numbers (0-9). All the signs can be translated real-time.	Real time sign language detection	Signs that are similar are misinterpreted and causes decrease in accuracy, smaller database
#4	Classifier is trained on both positive and negative data, weak classifier with lowest error rate is chosen in each run and then all the classifiers are combined as a strong classifier	The results show that the proposed AdaBoost system outperforms traditional AdaBoost	Position detecting rate increases, proposed hand detector improves false detecting rates, Proposed method can detect hand signs in bad background ,extracts accurate hand size and classifies left or right hand	False detecting Rate also rises because arm parts are segmented as a skin color.
#5	Vision based –> sign extraction from segmented B&W image and compare with database. Glove based -> Flex sensor output of glove sent to fuzzy logic system and then to microcontroller and RF transreceiver to display output.	The systems achieved 94% accuracy for Arabic Sign Language and 95% accuracy for American Sign Language.	Lower cost, can be compacted in mobile unit, not affected by outer environment, one glove is used for both English and Arabic languages. Vision based approach is easy.	Vision system is dependent on camera alignment and environment ambience.
#6	Frame capture and skin detection, hand segmentation and selecting features, reduction to 50x50 image, reduce dimensionality using LPP and classify using SVM.	Accuracy when trained on local machine is 75.29% whereas when done on cloud was 75.34%.	24% reduction in classification time, 32% less memory and 43% in processing time when assisted by cloud.	Power usage more by 3% when using 3G networks for communication with cloud. Database can be improved.
#7	Flex sensor, tactile sensor and accelerator based gesture recognition module for detecting hand gesture, convert output text to speech using Hidden Markov Model.	Average gesture recognition for A-H is 80-90% whereas it is 87.5% for the whole dataset.	Works for few common words also, accuracy is high, requires few components only, consumes low power, portable.	Implementation is not done for phrases and sentences.
#8				

Ref. No	**Objective**
#9	Develop computer based intelligent system to enable communication in SL, count the number of fingers opened in gesture of Binary SL.
#10	To build a 3D CNN to extract temporal features from video stream
#11	Mobile vision-based sign language translation device for automatic translation of Indian sign language into speech in English
#12	Detecting and tracking hand signs in a video using a new training method for Haar-like features based AdaBoost classifier
#13	Develop vision based and wireless glove based system for translating American and Arabic Sign Language using fuzzy logic procedure.
#14	Develop a mobile application to translate sign to speech using classification algorithms.
#15	Develop a system to recognize and translate sign language to bridge communication gap.

4 Conclusion

Sign language is a visual language. Visual information is the most important type of information perceived, processed and interpreted by the human brain. Digital image processing, as a computer-based technology has applications in a variety of fields such as image sharpening, restoration, medical, remote sensing etc. Also, deep learning, as a sub field of machine learning tries to imitate the actions of the human brain and is used in fields like speech and image recognition.

After going through the above listed papers on sign language conversion, we feel it is safe to say that there are many techniques used to convert sign language to various output types each with its own advantages and drawbacks in some. Our plan is to develop a mobile application that helps convert sign language to text output.

Acknowledgment

The authors express gratitude towards the assistance provided by our mentors and faculty members who guided us throughout the research and helped us in achieving desired results.

References

1. Kshitij Bantupalli, Ying Xie: “American Sign Language Recognition using Deep Learning and Computer Vision”, 2018 IEEE Conference on Big Data.
2. Krishna Modi, Amrita More: “Translation of Sign Language Finger - Spelling to Text using Image Processing”, International Journal of Computer Applications, 11 September 2013, vol. 77.
3. Dalal Abdulla, Shahrazad Abdulla, Rameesa Manaf, Anwar H. Jarndal: “Design and Implementation of A Sign to Speech/Text System for Deaf and Dumb People”, 2016 Fifth International Conference on Electronic Devices, Systems and Applications

(ICEDSA).

4. Kusumika Krori Dutta, Satheesh Kumar Raju K, Anil Kumar G S, Sunny Arokia Swamy B: “Double Handed Indian Sign Language to Speech and Text”, 2015 Third International Conference on Image Information Processing.
5. Padmavathi. S, Saipreethy M S, Valliammai V: “Indian Sign Language character recognition using Neural Networks”, IJCA Special Issue on Recent Trends in Pattern Recognition and Image Analysis RTPRIA.
6. M Suresh Anand, A. Kumaresan, Dr. N Mohan Kumar: “An integrated two way ISL(Indian Sign Language) translation system - A new approach”, International Journal of Advanced Research in Computer Science, Jan/Feb2013, Vol. 4 Issue 1, p7-12. 6p.
7. Eng-Jon Ong, Richard Bowden: “A Boosted Classifier tree for Hand Shape Detection”, Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.
8. Sachin Bhat, Amruthesh M, Ashik Chidanandas, Sujith: “Translating Indian Sign Language to text and voice messages using flex sensors”, International Journal of Advanced Research in Computer and Communication Engineering, May 2015, Vol. 4 Issue 5.
9. Sawaant Pramada, Deshpande Saylee, Naale Pranita, Nerkar Samiksha, Mrs. Archana S Vaidya: “Intelligent Sign Language recognition using Image Processing”, IOSR Journal of Engineering, Feb 2013, Vol. 3 Issue 2, pp 45-51.
10. Jie Huang, Wengang Zhou, Houqiang Li, Weiping Li: “Sign Language Recognition using 3D Convolutional Neural Networks”, 2015 IEEE Conference on Multimedia and Expo (ICME), 1-6, 2015.
11. Yellapu Madhuri, Anitha G, Anburajan M: “Vision-based Sign Language Translation Device”, 2013 International Conference on Information Communication and Embedded Systems (ICICES), 565-568, 2013.
12. Shuqiong Wu, Hiroshi Nagahashi: “Real-time 2D hands detection and tracking for Sign Language Recognition”, Proceedings of the 2013 8th International Conference on System of Systems Engineering, Maui, Hawaii, USA - Jun 2-6, 2013.
13. Anwar Jarndal, Ahmed Al-Maflehi: “On Design and Implementation of A Sign-to-Speech/Text System”, 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).
14. Philip Hays, Raymond Ptucha, Roy Melton: “Mobile Device to Cloud co-processing of ASL Finger Spelling to Text Conversion”, 2013 IEEE Western New York Image Processing Workshop (WNYIPW), 22-23 Nov, 2013.
15. Vijayalakshmi P, Aarthi M: “Sign Language to Speech Conversion”, 2016 International Conference on Recent Trends in Information Technology, 8-9 April, 2016.